Validity and reliability of critical power field testing

Karsten, B.; Jobson, S. A.; Hopker, J.; Stevens, L.; Beedie, C

Published in:
European Journal of Applied Physiology

DOI:
10.1007/s00421-014-3001-z

Publication date:
2015

Citation for published version (APA):
Karsten, B., Jobson, S. A., Hopker, J., Stevens, L., & Beedie, C. (2015). Validity and reliability of critical power field testing. *European Journal of Applied Physiology*, *115*(1), 197-204. https://doi.org/10.1007/s00421-014-3001-z

This is the Author's Accepted Manuscript version of a paper published in EUROPEAN JOURNAL OF APPLIED PHYSIOLOGY by Springer Berlin Heidelberg on 27 September 2014. The final publication is available at Springer via http://dx.doi.org/10.1007/s00421-014-3001-z.

Title: Validity and Reliability of Critical Power Field Testing

Running Title: Laboratory versus Field Critical Power Testing in Cycling

---

Karsten, B.[1], Jobson, S.A.[2], Hopker, J.[3], Stevens, L.[1], Beedie, C.[4]

[1]**School of Science,**
University of Greenwich,
Kent, ENGLAND
[2]**Department of Sport Studies,**
University of Winchester,
Hampshire, ENGLAND
[3] **School of Sport and Exercise Sciences,**
University of Kent,
Kent, ENGLAND
[4] **Department of Sport and Exercise Science,**
Aberystwyth University,
Wales, ENGLAND

✉ **Address for Correspondence:**
Bettina Karsten
School of Science,
University of Greenwich,
Kent, ME4 4TB
ENGLAND
Tel: +44 208 331 7927 – Fax: +44 208 331 9805
E-mail: B.Karsten@greenwich.ac.uk

Abstract

Purpose: To test the validity and reliability of field critical power (CP). **Method:** Laboratory CP tests comprised of three exhaustive trials at intensities of 80%, 100% and 105% maximal aerobic power and CP results were compared with those determined from the field. Experiment 1: cyclists performed three CP field tests which comprised maximal efforts of 12 min, 7 min and 3 min with a 30 min recovery between efforts. Experiment 2: cyclists performed 3 x 3 min, 3 x 7 min and 3 x 12 min individual maximal efforts in a randomised order in the field. Experiment 3: the highest 3 min, 7 min and 12 min power outputs were extracted from field training and racing data. **Results:** Standard error of the estimate of CP was 4.5%, 5.8% and 5.2% for experiments 1–3 respectively. Limits of Agreement for CP were -26 – 29 W, 26 – 53 W and – 34 – 44 W for experiments 1–3 respectively. Mean coefficient of variation in field CP was 2.4%, 6.5% and 3.5 % for experiments 1–3 respectively. Intraclass correlation coefficients of the three repeated trials for CP were 0.99, 0.96 and 0.99 for experiments 1–3 respectively. **Conclusions:** Results suggest field-testing using the different protocols from this research study, produce both valid and reliable CP values.

Key Words: Critical power; critical velocity; field testing capacity; reliability; validity.

Abbreviations

CoV	Coefficient of Variation (%)
CP	Critical Power
ICC	Intraclass Correlation Coefficient
LoA	Limits of Agreement
MAP	Maximal Aerobic Power (W)
PO	Power Output (W)
SEE	Standard Error of the Estimate
TTE	Time to Exhaustion
VO_{2max}	Maximal Aerobic Capacity (mL·min^{-1})

Introduction

Critical Power (CP), defined as the highest sustainable rate of aerobic metabolism (Gaesser and Wilson, 1988) without a continuous loss of homeostasis (Jones et al. 2008), demarcates the heavy and severe exercise domains (Poole 2009). CP furthermore provides an objective, valid, reliable, accurate and sensitive testing method (Jones et al. 2009) to monitor changes in endurance fitness (Stickland et al. 2000). However, commonly requiring multi-day testing, CP has not become a regularly assessed performance marker (Jones et al. 2010). This is an important research area, as effectively monitoring performances changes commonly seen in athletes requires valid and reliable tests. These tests have to be performed regularly to ensure that the training performed is achieving targeted adaptations. Recently, we have demonstrated that CP has a field application, by comparing CP obtained in the laboratory with that from testing in a velodrome and found low standard error of estimates (SEE; 2.5%), and a good level of agreement (LoA; -13.88–17.3 W) between the two environments (Karsten et al. 2013). Conversely, in agreement with previous literature (Green 1994), W′ (defined as an athlete's ability to exercise under increasing levels of fatigue caused by its own utilisation (Ferguson et al. 2010)) resulted in low LoA and high SEE values between the laboratory and field testing protocol (Karsten et al. 2014).

As our study was conducted on a cycling velodrome (Karsten et al. 2014) it did not indicate whether the agreement between CP values also holds true for road cycling, where the terrain can be flat or undulated. Nimmerichter et al. (2010) stated that PO is independent of external conditions and therefore potentially offers a more appropriate testing variable when designing field testing protocols. To be of use in all cycling events, any approach to the measurement of PO must be sufficiently sensitive to reliably detect small changes in PO that occur in the well trained athlete (Passfield et al. 2009). However, many of the field tests that coaches use with athletes are not sufficiently sensitive or reliable to provide a valid estimate of training effects (MacDougall et al. 1991).

A significant contribution to this research area was made by Quod et al. (2010) who investigated differences in PO values produced in the laboratory with those produced during road races. The study utilised elite cyclists who were assessed in their maximal capacity to produce power over defined durations set at 6, 15, 60, 60, 240 and 600 s. The final three maximal efforts (i.e. 60, 240 and 600 s) were also used to model CP and W′. Laboratory and field CP and W′ results were analysed and did not reveal a significant difference between the different environments. Whilst providing support for the validity of CP road testing, the study did not report the reliability of field CP values. If field-testing of CP is to be considered as a suitable and useful testing tool its reliability requires determination.

To our knowledge no study has attempted to validate a field test of CP from road cycling using a standardised testing protocol. Therefore the purpose of the present study was to compare CP determined in the laboratory with that modelled from maximal road effort durations of 12 min, 7 min and 3 min. The study further aimed to compare CP obtained from the highest 12, 7 and 3 minute power outputs recorded during a five week training period, with that from the laboratory. Finally the reliability of each respective CP field testing method was investigated. We hypothesised that there would be good agreement in CP calculated from laboratory and road exhaustive trials but not in W′. Further, we also hypothesised that both CP and W′ would demonstrate a good level of reliability –over three repeated field trials.

Methods

Participants in this study were recreational competitive road cyclists with a minimum of two years racing experience [minimum of 250–300 km or 10 h training volume per week]. The study was approved by the University Ethics Committee of the host institution. Prior to providing written informed consent, cyclists were fully informed of the nature and risks of the study. Participants were informed of their results on completion of the study. Eleven moderately trained cyclists (mean ± SD: age 32 ± 8 years, body mass 76.9 ± 14.9 kg, maximal aerobic power (MAP) 351 ± 37 W, maximal aerobic consumption ($\dot{V}O_{2\,max}$) 51.4 ± 9.8 $mL \cdot kg^{-1} \cdot min^{-1}$) participated in experiment one. Due to one drop out, 10 participants completed experiments two and three (mean ± SD: age 32 ± 9 yr, body mass 75.3 ± 15.1 kg, MAP 346 ± 36 W, $\dot{V}O_{2\,max}$ 51.9 ± 10.3 $mL \cdot kg^{-1} \cdot min^{-1}$).

Study design. Participants' road bikes were equipped with a PowerTap Elite wheel (CycleOps, Madison, USA) and a magnet for direct cadence measurement (Gardner et al. 2004). For the laboratory tests, the same road bike was attached to a Computrainer (RacerMate, Seattle, USA). The PowerTap device was zero offset prior to all trials according to the manufacturer's instructions (Gardner et al. 2004). During two visits to the laboratory, $\dot{V}O_{2\,max}$, MAP values and laboratory CP and W′ were determined. Participants refrained from heavy exercise in the 24 h prior to tests, and from food intake in the 3 h prior to tests. For both laboratory visits participants were instructed to arrive at the same time of the day. The field study contained three experimental protocols which are detailed below.

Maximal oxygen uptake test. Following a standardised warm-up, cyclists completed a progressive, incremental laboratory exercise test to exhaustion. The maximal test commenced at a work rate of 150 W. Thereafter, intensity increased at a step rate of 20 $W \cdot min^{-1}$ using power values obtained from the PowerTap. Cyclists were allowed to self-select their cadence and were instructed to maintain this cadence throughout the test, whilst remaining in the same cycling gear. When cadence dropped by more than 10 $rev \cdot min^{-1}$ for more than 10 s despite strong verbal encouragement, tests were terminated. Expired gases were collected continuously throughout the test using a Cortex MetaLyzer 3B gas analyser (Cortex Biophysik, Leipzig, Germany). Fingertip blood lactate was analysed using the Biosen C_line analyser (EFK Diagnostics, Barleben, Germany), and heart rate (HR) was continuously monitored using the Cortex gas analyser to ensure a maximal exhaustive test. MAP was calculated as the highest 30-s mean PO value (W) and $\dot{V}O_{2\,max}$ was calculated as the highest mean oxygen consumption over the same period.

Critical Power laboratory test. Participants completed three time to exhaustion (TTE) trials on the equipment described above. Work rates were equivalent to ~80%, ~100% and ~105% MAP, using a lowest to highest work rate order with a 30 min intra-trial recovery period (Galbraith et al. 2014). Unpublished data from our laboratory supports the use of the testing methods by Galbraith et al. in cycling. During rest periods fluid intake was permitted ad libitum. During each TTE trial, participants were cooled using an electric fan. Laboratory conditions were stable in a range of 18–22 C$^{\circ}$ with 45–55% humidity. After a 5-min warm-up at a work rate of 150 W, the test resistance was set and cyclists were instructed to maintain their preferred cadence for as long as possible. Tests were terminated when cadence, despite strong verbal encouragement, dropped by more than 10 $rev \cdot min^{-1}$ for more than 10 s. HR, PO and cadence were recorded continuously via the PowerTap, and expired gases were continuously sampled through the gas analyser to ensure the attainment of individual VO2max values. Strong verbal encouragement was provided throughout the tests whilst participants were blinded to TTE trial intensities and elapsed time. Fingertip capillary blood samples were taken prior to TTE trials and at test termination. All cyclists reached their individual $\dot{V}O_{2\,max}$ value (± 0.09 L·min^{-1}), a post-test blood [lactate] of ≥ 8 mmol·l^{-1} and a HR within ± 5 beats of their maximal HR values established during the $\dot{V}O_{2\,max}$ test.

Critical Power field tests. Within the racing season and over the duration of 5 weeks, cyclists were required to record their training and racing activities using the PowerTap. Participants were instructed to avoid freewheeling during 'purposeful' efforts. To achieve a high level of ecological validity, environmental conditions were not standardized and no instructions for the choice of road, gradient or cycling position were given. Cyclists were instructed to perform one unsupervised familiarization trial of experiment 1 and experiment 2, which were not included in the data.

Experiment 1 (N = 11);

CP and W′ were determined using 3 field-based tests. These comprised of a 12 min, followed by a 7 min and a final 3 min maximal effort using a recovery period of 30 min. Between maximal efforts cyclists either rested passively or continued cycling at a low, i.e. recovery intensity. Experiment 1 consequently resulted in three CP and three W′ values. Cyclists were instructed to perform these series of maximal efforts fully rested. Experiment 1 contains 9 purposeful efforts with a minimum of a 24 h recovery period between each set of maximal efforts.

Experiment 2 (N = 10);

CP and W′ were determined using 3 field-based tests, which were performed individually during single but randomised training sessions. Participants in this experiment in total had to complete three sets of required efforts of each 12 min, 7 min and 3 min maximal efforts over 9 individual training sessions. Cyclists were instructed to perform any of these maximal efforts fully rested. The completion of one set, i.e. a 3 min, a 7 min and a 12 min effort were used in the CP and W′ modelling process and experiment 2 consequently resulted in three CP and three W′ values. Experiment 2 contains 9 purposeful efforts with a minimum of a 24 h recovery period between individual efforts.

Experiment 3 (N = 10);

As some of the intentional efforts were lower than 'non-intentional' efforts, experiment 3 used the highest three PO values (12, 7 and 3 minute durations) of all training and racing files for the determination of CP and W′. Experiment 3 consequently resulted in three CP and three W′ values.

Calculation of Critical Power and W′. Training sessions were recorded via a Garmin Edge 500 head unit (Garmin International, Kansas, USA). Files were imported into WKO training software (Peaksware LLC, v3+, Boulder, USA) and the specified efforts (i.e. 12 min, 7 min and 3 min) extracted to model CP and W′ for experiments 1, 2 and 3. For all experiments, linear regression was used to determine CP and W′ using the power-1/time (P = W′(1/t) + CP model. Results determined from individual experiments were consequently termed CP1/CP2/CP3 and W′1/W′2/W′3.

Statistical Analysis. Data were first examined using the Shapiro-Wilk normality test. Both, the validity and the reliability of field CP and W′ values were assessed within each experiment. To assess the variability of results from experiments 1–3, the within subject variation, expressed as a Coefficient of Variation (CoV) and Intra Class Correlation (ICC) were used. Repeated measures

ANOVA was used to test for significant differences between repeated trials. Pearson product moment correlation analysis was used to provide an indication of the strength of any relationship between the laboratory values for CP and W′ and the different field test values. Agreement between the laboratory values and all mean experimental field values of CP and W′ was assessed using LoA (Atkinson and Nevill 1998; Bland and Altman 1986). Linear regression was used to calculate values for SEE for CP and W′ in each experiment (as mean values) and for laboratory values of CP and W′. Differences of statistical significance between laboratory and mean field values of CP and W′ were tested using paired samples t-tests and accepted at $P < 0.05$. Results are reported as mean ± SD unless otherwise stated.

RESULTS

Agreement between lab and field CP and W'. Laboratory CP values was significantly correlated with experimental field CP values ($P \leq 0.01$). Laboratory TTE trials durations were 667 ± 176 s, 256 ± 105 s, and 143 ± 44 s at 80%, 100% and 105% MAP respectively. The paired samples t-tests did not reveal any significant differences between laboratory and field CP values for all experiments ($P = > 0.05$). Significant differences ($P = < 0.05$) between laboratory and field values of W′ were demonstrated for experiment 2 and experiment 3. LoA and SEE values for all experimental CP and W′ values are presented in **table 1a & b** respectively**,** with **figures 1 - 3** illustrating Bland-Altman plots of laboratory and mean field values of CP for all experiments.

Table 1a and 1b about here

Figure 1 about here

Figure 2 about here

Figure 3 about here

Reliability of experimental protocols. For all experiments, repeated measures ANOVA identified no significant differences (i.e. bias) in CP between trials within each experiment. (Experiment 1, CP, $F(2, 9) = 1.64$, $P > .05$. Experiment 2, CP, $F(2, 8) = 0.20$, $P > .05$. Experiment 3, CP, $F(2, 8) = 3.33$, $P > .05$). CoV values for experiment 1 ranged between 2.2% and 2.5%, for experiment 2 the range was between 5.9% and 7.0% and for experiment 3 it was between 3.3% and 3.6% **(Table 2a).** Intraclass correlation coefficient for all experiment protocols for CP ranged between 0.96 and 0.99 (95% CI 0.90 – 0.99) (**Table 2a**). Mean laboratory SEE values for CP resulted in 5 ± 3.07 W.

Table 2a and 2b about here

Discussion

The main findings of this study were a good level of agreement between laboratory and field determined values of CP for all experimental protocols. Furthermore laboratory CP strongly correlated with field CP and experimental CP field testing protocols generally had a very high test-retest reproducibility **(table 2a). Table 1a** demonstrates low mean, non-significant differences between field and laboratory CP values, acceptable LoA (Bland and Altman 1986) and low SEE values. Gonzales-Haro et al. (2007) accepted their incremental velodrome field test as being valid with reported LoA of 130 W to -24 W and a random error of 77.1 W (13.9%). The present study demonstrates LoA values which are considerably higher and SEE's that are considerably lower than those reported by Gonzales-Haro et al. (2007). Our recent study (Karsten et al. 2013) reported similar mean differences of 2 ± 8 W with LoA between 11 W and 17 W and SEE values of 2.5% to those in this current study when comparing CP determined in the laboratory with CP determined from the track. We therefore suggest that the experimental protocols can be considered to be acceptable when testing CP in the field. In particular the field test protocol used in experiment 1 provided the best agreement between laboratory and field CP values (**Fig. 1, panel A and B**). This is not surprising given an almost equal protocol in that CP testing was performed within a maximum testing duration of 2.5 hours, using the same order of maximal efforts and a 30 min recovery period between those efforts.

As hypothesised, low levels of agreement were found for field determined W′ values (**table 1b**). Moreover, experiment 2 and 3 identified significant differences between laboratory and field W′ with high prediction errors (≥ 29%) for all experimental values being evident. Previous research has questioned the reliability of W′ (Dekerle et al. 2006; Vandewalle et al. 1997). Although likely to be multifactorial, differences for W′ in experiment 1 might be due to differences in standing or rolling start, differences in power profiles between constant-load laboratory and time trial type field efforts (Karsten et al. 2014), or change of cadence with a change in terrain (Jobson 2008; Nimmerichter 2012)Adding to these influences and due to having performed relevant efforts on different days, experiment 2 and 3 might contain more environmental (for example changes in weather condition or humidity), time and circadian rhythm influences, which can impact on anaerobic power (Racinais et al. 2004). By contrast Quod et al. (2010), did not find any effect of location on W′ when comparing laboratory and race determined values. Moreover W′ in the present study appears to exhibit a lower test re-test reproducibility **(Table 2b)** which further compromises the validity of this parameter**.** Another issues to consider are that of ground level and gradient cycling and cycling position. Padilla et al. (1999) investigated differences between level and uphill time trials (TT) in professional cyclists. Mean PO was generally higher during uphill cycling and the authors suggested that higher PO can only be achieved during uphill cycling. This is further supported by Nimmerichter et al. (2012) who recently reported on low cadence up-hill 20-min uphill TT performances resulting in significantly higher PO, HR and blood lactate values when compared to 20-min flat TT performances. Investigating cycling positions, Jobson et al. (2008) demonstrated higher field PO values when performed in an aerodynamic position but there was no difference in PO values between field and the seated upright laboratory TTs. Most importantly, the study demonstrated the independence of PO from environmental conditions, as the higher mean road PO values were not reflected in higher velocities, when comparing the same cycling position. Given that no instructions were provided on how to perform the maximal efforts nor where to perform them, an undulated terrain and possible changes in cycling position might have contributed to the differences in W′ due to an increased portion of type II fibre recruitment and the resultant higher PO values associated with greater blood lacatate concentrations (Sjödin 1976; Tesch et al. 1978).

A CoV of 10% has been suggested as the criterion value commonly used to define an acceptable level of test reliability (Atkinson et al. 1999). To verify a reliable test Atkinson et al. further suggested an ICC > 0.8. Hopkins (2000) later defined a lower 5 % CoV as the acceptable upper limit in sports science reliability studies. Given that the CoV values for CP observed during experiment 1 and experiment 3 **(Table 2a)** were below the lower boundary as defined by Hopkins (2000) the respective experimental testing protocols can be deemed as being reliable. High interclass correlation coefficients (i.e. > 0.8; **table 2a**) further demonstrate the repeatability of all experimental protocols with a small bias ± random error, which are considerably lower than those reported by Gonzales-Haro et al. (2007). Experimental protocol 2 resulted in mean CoV values of 6.5% for CP which according to Atkinson et al. (1999) can also be deemed as acceptable. However, poorer LoA and higher associated prediction errors **(Table 1a)** means that it is reasonable to question whether protocol 2 is as good as protocols 1 and 3 in its ability to accurately monitor the small changes in CP typically seen in trained athletes (Passfield et al. 2009). Furthermore, our hypothesis of W′ demonstrating a good level of reliability across repeated field trials has to be rejected. CoV and ICC values for all experimental protocols were higher than the defined values by Atkinson et al. (1999) or Hopkins (2000) and it is questionable whether this parameter of the power-duration relationship is either valid or reliable in field testing.

The present study collected data over the duration of a 5-week period, towards the end of the racing season. Whilst assuming that CP would remain stable over this time period, small performance changes which may have affected results cannot be eliminated (Nimmerichter et al. 2011). Cyclists were required to conduct a total of 18 purposeful efforts of 12, 7 and 3 minute durations during the period. Attempting to have minimal impact on regular training, cyclists were not required to conduct the efforts in any order or at any specific time point. Interestingly, the results of the current study are supportive of our previous work conducted in an outdoor velodrome, where the cyclist were performing within a consistent and more predictable environment (Karsten et al. 2014). Using a similar approach as in experiment 2, cyclists had to perform maximal efforts of fixed durations of 12, 7 and 3 min on separate days, and in a randomised order. We found a high agreement for CP but not for W′ when comparing laboratory and velodrome environments. However reported values for LoA of CP in the present study **(table 1a, experiment 2)** are not as high as in our velodrome study (Karsten et al. 2013), which possibly demonstrates an influence of terrain on CP.

In experiment 3, we extracted the single highest 3 min, 7 min and 12 min efforts from all of the training and racing files. Cyclists were not given instructions as to where to perform or how to perform these maximal efforts (i.e. seated or standing). Whilst laboratory trials were solely performed in a seated position results demonstrated a high level of agreement with field CP values (LoA -34 – 44 W; SEE 14 W). Using a similar approach to the current study, Quod et al. (2010) extracted maximal efforts of fixed durations over 1 min, 4 min and 10 min to model CP and W′ from race data. In agreement with our findings **(table1a)**, Quod et al. (2010) did not find a significant difference between laboratory and field CP results. However, it has to be noted that the lowest duration of 60 s used by Quod et al. does not comply with the requirements of CP testing as set by DiPrampero (1999), i.e. attainment of $\dot{V}O_{2max}$. Furthermore the power profile testing, which included relevant CP and W′ efforts was not validated against conventional CP testing standards, as the researchers utilised an active recovery performed at 100 W for individual break periods (330 s, 480 s and 600 s between relevant maximal efforts of 60 s, 240 s and 600 s respectively). Interestingly, our data demonstrate a trend for higher mean field PO's in experiment 3, compared to those of experiment 1, 2 and in the laboratory. Training files revealed, on a number of occasions, that cyclists produced higher mean PO of the set duration efforts during experiment 3, i.e. during efforts extracted from regular training and racing data. It can be suggested, that unintentional efforts might have minimised the impact of pacing on resultant mean PO about values. In short, having not been aware of being tested, cyclists did not apply any clear pacing strategy during relevant efforts. However, the higher mean PO values did not appear to greatly influenced values of CP, just W′. Deemed as being reliable (mean CoV 3.5%; ICC 0.99), the protocol used in experiment 3 could therefore provide a valid other method of assessing CP from 'normal' training efforts during which the cyclist does not have to provide pre-defined 'intentional' efforts.

This is the first study to demonstrate that CP can be determined in the field under 'controlled' (i.e. planned maximal efforts for a given protocol) and 'uncontrolled' (i.e. extraction of data from training and performances) situations. In particular the protocol used in experiment 1 resulted in a high level of agreement (-2 ± 14 W) and low prediction errors (< 5%), whilst providing a more ecologically valid testing environment when compared to laboratory testing. When applying experimental protocol 2 and 3, lower LoA values and higher prediction errors have to be acknowledged but in spite of this, both protocol 2 and 3 have the advantage of being more easily integrated into the training schedule of riders. Each proposed CP field testing protocols can therefore be recommended to coaches and athletes as routine assessment. Future research studies are recommended to analyse training related changes in CP throughout the racing season, in particular applying field-testing protocols 1 and 3, which provided the lowest CoV values.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the cooperation of the athletes who took part in this study. No funding was received for this work and there is no conflict of interest.

References

Atkinson, G., Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med (Auckland, N.Z.), 26(4), 217–238

Atkinson G., Nevill A.M., Edwards. B. (1999). What is an acceptable amount of measurement error? The application of meaningful “analytical goals” to the reliability analysis of sports science measurements made on a ratio scale. J Sports Sci 117, 18

Bland, J. M., Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1(8476), 307–310

Dekerle, J., Brickley, G., Hammond, A J. P., Pringle, J. S. M., Carter, H. (2006). Validity of the two-parameter model in estimating the anaerobic work capacity. Eur J Appl. Physiol 96(3), 257–264

DiPrampero, P. E. (1999). The concept of critical velocity: a brief analysis. Eur J Appl. Physiol. Occup Phys 80(2), 162–164

Ferguson, C., Rossiter, H. B., Whipp, B. J., Cathcart, A. J., Murgatroyd, S. R., Ward, S. A. (2010). Effect of recovery duration from prior exhaustive exercise on the parameters of the power-duration relationship. J Appl. Physiol *(Bethesda, Md. : 1985)*, 108(4), 866–874

Gaesser, G. A., Wilson, L. A. (1988). Effects of continuous and interval training on the parameters of the power-endurance time relationship for high-intensity exercise. Int J Sport Med 9(6), 417–421

Galbraith, A., Hopker, J., Lelliott, S., Diddams, L., Passfield, L. (2014). A Single-Visit Field Test of Critical Speed. Int J Sports Physiol Perform doi:10.1123/ijspp.2013-0507

Gardner, A. S., Stephens, S., Martin, D. T., Lawton, E., Lee, H., Jenkins, D. (2004). Accuracy of SRM and Power Tap Power Monitoring Systems for Bicycling. Med Sci Sports Ex 36(7), 1252–1258

González-Haro, C., Galilea, P. A., Drobnic, F., Escanero, J. F. (2007). Validation of a field test to determine the maximal aerobic power in triathletes and endurance cyclists. Brit J Sport Med 41(3), 174–179

Green, S. (1994). A definition and systems view of anaerobic capacity. Eur J Appl. Physiol. Occup Phys 69(2), 168–173

Hopkins, W. G. (2000). A new view on statistics. Internet Society for Sport Science. Available at: from http://www.sportsci.org/resource/stats/. Accessed 07.05.2014

Jobson, S. A., Nevill, A. M., George, S. R., Jeukendrup, A. E., & Passfield, L. (2008). Influence of body position when considering the ecological validity of laboratory time-trial cycling performance. J Sports Sci, 26(12), 1269–1278

Jobson, S. A., Woodside, J., Passfield, L., Nevill, A. M. (2008). Allometric scaling of uphill cycling performance. Int J Sport Med 29(9), 753–757

Jones, A. M., Vanhatalo, A., Burnley, M., Morton, R. H., Poole, D. C. (2010). Critical power: implications for determination of VO2max and exercise tolerance. Med Sci Sports Exerc 42(10), 1876–1890

Jones, A. M., Vanhatalo, A., Doust, J. (2009). Kinanthropometry and Exercise Physiology Laboratory Manual: Exercise physiology. (Eston & Reilly, Eds.) (pp. 271–306). London, UK: Taylor & Francis

Jones, A. M., Wilkerson, D. P., DiMenna, F., Fulford, J., Poole, D. C. (2008). Muscle metabolic responses to exercise above and below the “critical power” assessed using 31P-MRS. *Am J Physiol Regul Integr Comp Physiol* 294(2), R585–593

Karsten, B., Jobson, S. A, Hopker, J., Jimenez, A, Beedie, C. (2013). High Agreement between Laboratory and Field Estimates of Critical Power in Cycling. Int J Sports Med 35(4), 298–303

MacDougall, J.Duncan; Wenger HA, Green, H. (1991). Physiological Testing of the High-performance Athlete (2nd ed.). Champaign, IL: Human Kinetics Publishers.

Nimmerichter, A., Eston, R., Bachl, N., & Williams, C. (2012). Effects of low and high cadence interval training on power output in flat and uphill cycling time-trials. Eur J Appl Physiol 112(1), 69–78

Nimmerichter, A., Eston, R. G., Bachl, N., Williams, C. (2011). Longitudinal monitoring of power output and heart rate profiles in elite cyclists. J Sports Sci 29(8), 831–40

Nimmerichter, A., Williams, C., Bachl, N., Eston, R. (2010). Evaluation of a field test to assess performance in elite cyclists. Int J Sport Med 31(3), 160–166

Padilla, S., Mujika, I., Cuesta, G., Goiriena, J. J. (1999). Level ground and uphill cycling ability in professional road cycling. Med Sci Sports Exerc 31(6), 878–885

Passfield, L., Jobson, S. A., Atkinson, G., Barton, G., Scarf, P. (2009). Longitudinal changes in weighted indices of work-rate but not average power output of cyclists during a racing season. J Sports Sci 27(S2), S102–103

Poole, D. C. (2009). Resolving the determinants of high-intensity exercise performance. *Exp Physiol* 94(2), 197–198

Quod, M., Martin, D., Martin, J. (2010). The power profile predicts road cycling MMP. Int J Sports Med 31(6), 397–401

Racinais, S., Hue, O., Blonc, S. (2004). Time-of-day effects on anaerobic muscular power in a moderately warm environment. *Chronobiol Int* 21(3), 485–495

Sjödin, B. (1976). Lactate dehydrogenase in human skeletal muscle. *Acta Physiol* Scand Suppl 436, 1–32

Stickland, M. K., Petersen, S. R., Dressendorfer, R. H. (2000). Critical aerobic power during simulated 20 km bicycle racing. Sports Med Training Rehab 9(4), 289–301

Tesch, P., Sjödin, B., Karlsson, J. (1978). Relationship between lactate accumulation, LDH activity, LDH isozyme and fibre type distribution in human skeletal muscle. *Acta Physiol* Scand 103(1), 40–46

Vandewalle, H., Vautier, J. F., Kachouri, M., Lechevalier, J. M., Monod, H. (1997). Work-exhaustion time relationships and the critical power concept. A critical review. J Sports Med Phys Fitness 37(2), 89–102

Tables and Figures

Table 1a. Mean values, mean differences, limits of agreement and standard error of estimate of CP. Values are mean (± SD)

	Experiment 1	Experiment 2	Experiment 3
Mean Field CP (W)	277 ± 38	271 ± 44	276 ± 46
Mean Difference CP lab (W)	-2 ± 14	10.37 ± 22	-5 ± 20
95% CI	-11.19 - 7.74	-26.06 - 5.06	-19.31 - 9.31
LoA (W)	-26 – 29	-32 - 53	-34 - 44
SEE (%)	4.5	5.8	5.2
SEE (W)	11	17	14

Table 1b. Mean values, mean differences, limits of agreement and standard error of estimate of W'. Values are mean (± SD) - * = significantly difference from laboratory W' values.

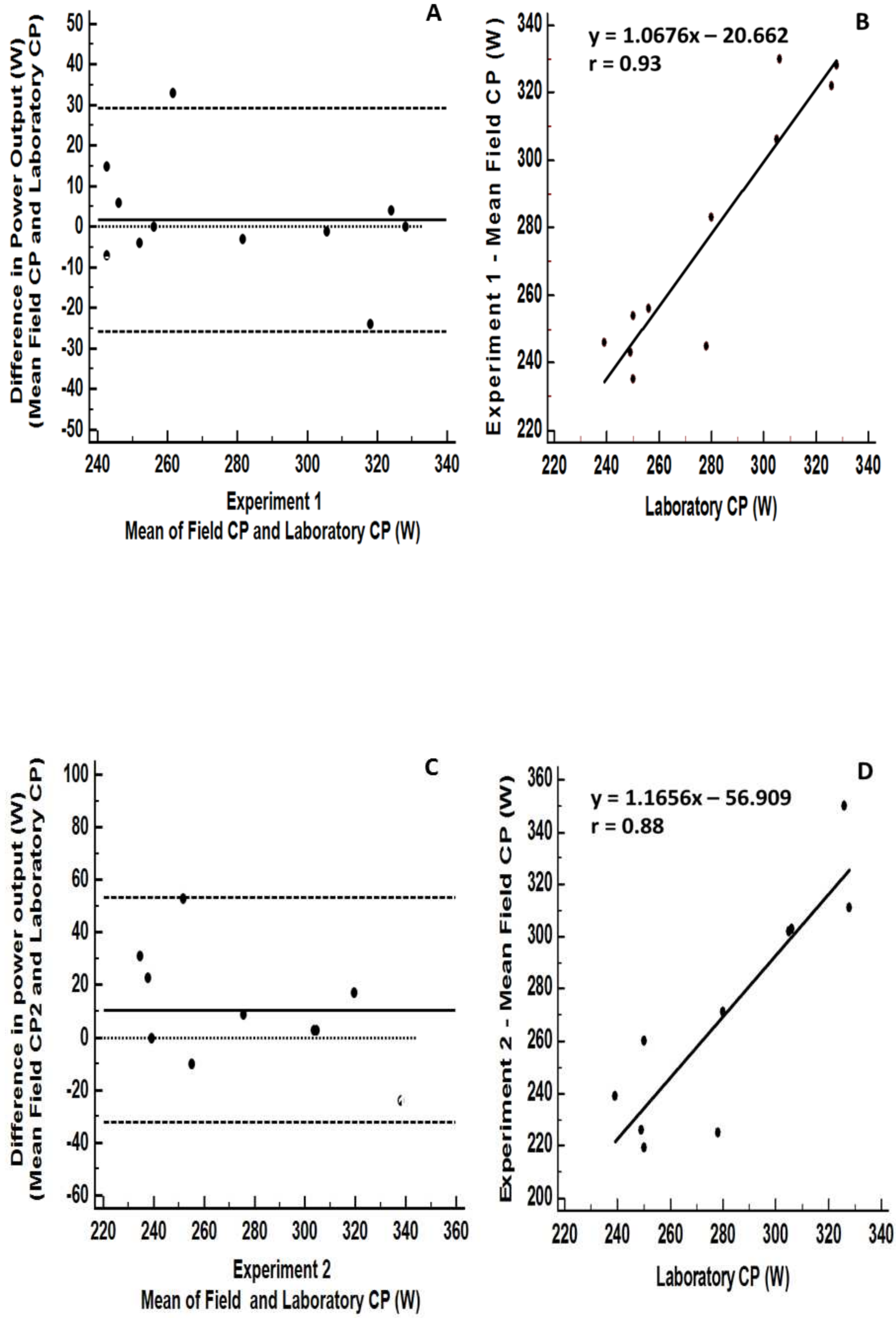
	Experiment 1	Experiment 2	Experiment 3
Mean Field W' (kJ)	12 ± 3	17 ± 5	20 ± 5
Mean Difference W' lab (kJ)	-0.14 ± 3.36	-4.62 ± 5.69*	7.79 ± 3.15*
95% CI	2.40- 2.12	0.54 - 8.69	5.53 - 10.04
LoA (kJ)	-6 - 7	-16 - 7	-14 to -2
SEE (%)	31.4	39.4	31.8
SEE (kJ)	3.08	4.03	2.83

Table 2a. Coefficient of Variations (CoV) values, Intraclass Correlation Coefficient (ICC) values and 95% Confidence Intervals (CI) for all experimental CP results

Field Tests	Experiment 1 CP (W)	Experiment 2 CP (W)	Experiment 3 CP2 (W)
CoV (%) Trial 1 vs trial 2	2.5	7.0	3.6
CoV (%) Trial 2 vs trial 3	2.2	5.9	3.3
ICC	0.99	0.96	0.99
95% CI	0.98 - 0.99	0.90 - 0.99	0.96 - 0.99

Table 2b. Coefficient of Variations (CV) values, Intraclass Correlation Coefficient (ICC) values and 95% Confidence Intervals (CI) for all experimental W′ results

Field Tests	Experiment 1 W′ (kJ)	Experiment 2 W′ (kJ)	Experiment 3 W′ (kJ)
CoV (%) Trial 1 vs trial 2	46.7	48.3	15.6
CoV (%) Trial 2 vs trial 3	46.0	41.5	17.9
ICC	0.16	0.028	0.63
95% CI	-0.82 – 0.81	-0.29 – 0.44	0.23 – 0.89


A
Difference in Power Output (W)
(Mean Field CP and Laboratory CP)
Experiment 1
Mean of Field CP and Laboratory CP (W)
B
y = 1.0676x – 20.662
r = 0.93
Experiment 1 - Mean Field CP (W)
Laboratory CP (W)
C
Difference in power output (W)
(Mean Field CP2 and Laboratory CP)
Experiment 2
Mean of Field and Laboratory CP (W)
D
y = 1.1656x – 56.909
r = 0.88
Experiment 2 - Mean Field CP (W)
Laboratory CP (W)

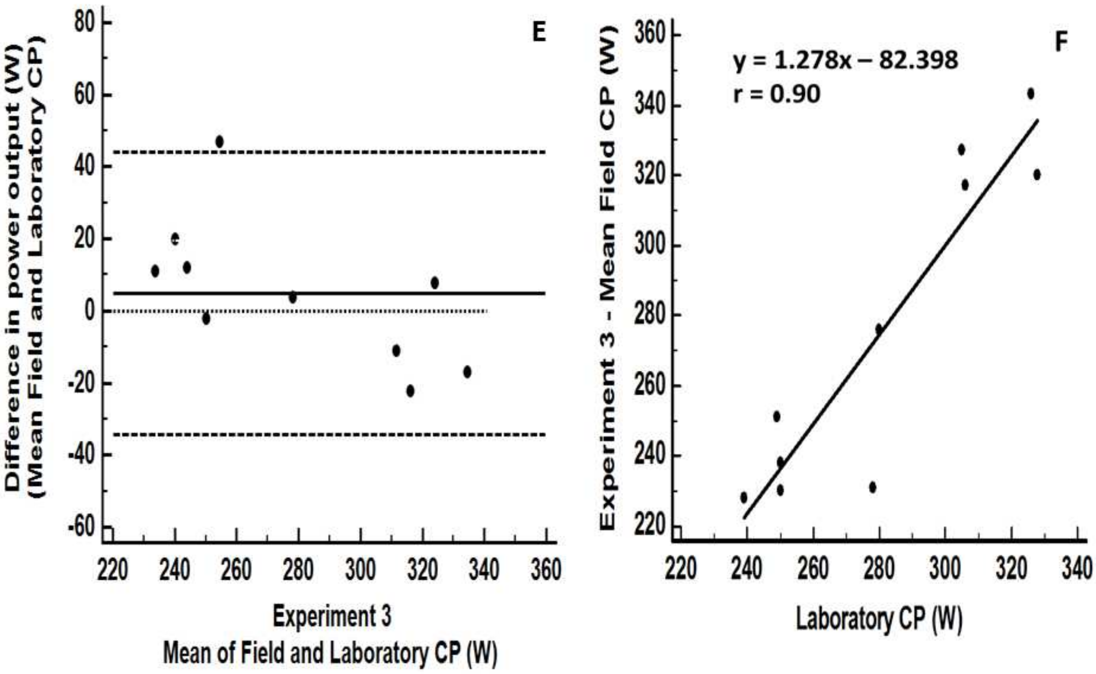


Figure captions

Fig. 1-3. Bland-Altman plots of the limits of agreement (panel A, C and E) and the relationship (panel B, D and F) between laboratory CP and field CP (W). In panel A, C, and E the horizontal line represent the mean difference between laboratory CP and field CP, and the dashed line represents 95% LoA.